

Kirjojen skannaus

Eryteisesti vanhojen tekijänoikeusvapaiden kirjojen skannaus on hyödyllistä, sillä silloin ne voi jakaa helposti muille ilman että tarvitsee pelätä kirjan kunnan puolesta. Tekijänoikeuden suojaamista kirjoistakin saa tehdä digitaalisia kopioita, mutta vain yksityiseen käyttöön.

□

Sisällys

- [1 Tekijänoikeuslaki](#)
- [2 Skannerit](#)
- [3 Ohjelmistot](#)
 - [3.1 Tekstintunnistusohjelmat](#)
- [4 Linkit](#)

Tekijänoikeuslaki

- Kun 70 vuotta tekijän kuolemasta on kulunut, kirjat ovat tekijänoikeuksista vapaita.

Skannerit

- [CanoScan LiDe 25](#) (55 euroa)

- *Linux - Mac OS X - Windows*

- Halvin skanneri, joka ainoana kaikista toimii kaikilla alustoilla. Käyttää led-tekniikkaa skannauksessa. Mukana tulee OCR-ohjelma Omnipage SE, joka ei kuitenkaan ole parhainta tasoa.

- [Ohje: asentaminen Ubuntu Linuxiin.](#)

Mikäli aikomuksena on digitoida useampia teoksia kannattaa harkita erityisen kirjaskannerin hankkimista. Kaupalliset skannerit ovat usein kalliita ja niiden saatavuus heikko, joten [tee-se-itse - kirjaskannerin](#) rakentaminen voi olla varteenotettava vaihtoehto. Digitaaliseen järjestelmäkameraan perustuvalla skannerilla voidaan voidaan päästä jopa yli 10 skannattuun sivuun minuutissa.

Ohjelmistot

- [Scan Tailor](#)

- *Windows, Linux*

- Ilmainen ohjelma skannattujen sivujen jälkikäsitteilyyn. Ohjelmalla voi mm. jakaa aukeamat sivuiksi, suoristaa vinoutuneet/vääntyneet sivut sekä poistaa ja lisätä reunuksia. Käsitellyt sivut soveltuvat hyvin suoraan PDF-tiedostoksi muuttamiseen tai tekstintunnistusohjelmalla käsiteltäviksi.

Tekstintunnistusohjelmat

Skannatut sivut eivät itse asiassa ole tekstiä vaan kuvia. Kuvatiedostojen käyttäminen sellaisenaan sähköisessä kirjassa ei ole mielekästä, sillä ne vievät huomattavasti enemmän tilaa kuin teksti eikä kirjan sisältöä voi muokata helposti. Tämä voi koitua ongelmalliseksi jos kirjaa halutaan lukea sähköisillä lukulaitteilla, sillä monet lukulaitteet eivät tue kuvista koostuvia PDF-tiedostoja. Lisäksi "kuvakirjan" fonttikokoa ja asettelua ei voi muuttaa.

Ratkaisu on tekstintunnistus (engl. OCR, Optical Character Recognition). Tekstintunnistusohjelmat nimensä mukaisesti tunnistavat kuvassa olevan tekstin ja tallentavat sen sitten esimerkiksi tekstinkäsittelydokumenttina tai suoraan PDF-tiedostoksi.

- [Abby FineReader 9](#)

- *Windows*

- Ilmeisesti tarkin ja siistein tunnistusohjelma. Sen takia myös suositeltavin, jotta siistimistyö jää mahdollisimman vähäiseksi. Toimii vain yhdellä käyttöjärjestelmällä, mutta esim. [VMwaren](#) tai [Virtualboxin](#) avulla voi sekä Linuxilla että Mac OS X:llä ajaa Windowsia.

- [Omnipage 16](#)

- *Mac OS X - Windows*

- Tätä ohjelmaa on paikoin kehuttu parhaaksi, eikä se huono olekaan sillä se on mm. nopein. Kuitenkin lopputuloksen siisteys jää toiseksi FineReaderille. [\[1\]](#)

- [ReadIris Pro 11](#)

- [SimpleOCR](#)

- *Windows*

- ILMAINEN. Sopii lähinnä muutaman sivun skannaukseen huonon tarkkuuden johdosta.

- [Tesseract](#)

- *Linux - Windows* (epävirallisesti myös *Mac OS X*)

- Avoimen lähdekoodin tekstintunnistusohjelma yli 40 kielelle. Ohjelma on komentorivipohjainen, mutta sille on graafisia käyttöliittymiä; mm. [OCRFeeder](#) Linuxille ja [FreeOCR](#) Windowsille.

Linkit

- [Wikipedia - OCR](#)

- [ORC Software Guide](#)

- [WPU Forums - Best ORC software?](#)

- [OCR - What It Is, How to Do it Better](#)

- [Helsingin kirjasto - Raportti aineiston muuttamisesta HTML-muotoon](#)